Wireless Sensor Networks for Intrusion Detection: Packet Traffic Modeling

Ilker Demirkol, *Student Member, IEEE,* Fatih Alagöz, *Member, IEEE,* Hakan Deliç, *Senior Member, IEEE,* and Cem Ersoy, *Senior Member, IEEE*

***Abstract*— Performance evaluation of wireless sensor network (WSN) protocols requires realistic data traffic models since most of the WSNs are application specific. In this letter, a sensor network packet traffic model is derived and analyzed for intrusion detection applications. Presented analytical work is also validated using simulations.**

***Index Terms*— Traffic model, wireless sensor networks, intrusion detection.**

I. Introduction

Wireless sensor networks (WSNs) have unique characteristics that separate them from legacy communication networks, and make it necessary to develop suitable traffic models. In the WSN literature, constant bit rate (CBR) data traffic is commonly employed (e.g., [1]). There are also a few studies that use variable bit rate (e.g., Poisson [2]) data sources. However, the traffic load generated heavily depends on the application, which can be categorized as *event-driven* or *periodic data generation*. The event-driven scenarios such as *target detection* and *tracking* generate bursty traffic which cannot be modeled as either CBR or Poisson. Therefore, one needs to define the set of parameters for the application in hand before proceeding with the traffic model development.

Constructing accurate and analytically tractable source models for sensor network traffic will provide a basis for further work on proposed network protocols. Performance evaluation of WSNs will be performed with realistic traffic loads. Besides, the effects of system parameters such as node density and target velocity can be analyzed without the need for simulations.

In this letter, we consider a new packet traffic model for sensor networks whose task is intrusion detection. For intrusion detection applications, the predetermined surveillance area (e.g., a border) is generally represented in the form of a square grid. In addition, possible sensor node locations are often defined to be the grid corners even if the deployment is random. However, the quality of this kind of modeling depends heavily on the distance between the two nearest grid cross-points, which must be infinitely small for the best simulation of real life. To overcome this deficiency, we do not try to specify the deployment locations of the sensor nodes, but instead, we define probabilistic coverage degrees of surveillance area points for the uniformly distributed sensor scenario. Thus, the requirement for any grid structure is eliminated. The probabilistic coverage degree model presented is similar to the probabilistic hidden terminal model used in [3] to calculate the probability of successful packet transmission. In Section II, we present this coverage model and in Section III, we analyze the packet traffic based on this coverage model. Finally in Section IV, simulation results are compared to the analysis.

Manuscript received June 14, 2005. The associate editor coordinating the review of this letter and approving it for publication was Prof. Enzo Baccarelli. This work is supported by the State Planning Organization of Turkey under the grant number 03K120250.

I. Demirkol, F. Alagöz, and C. Ersoy are with the Dept. of Computer Engineering, Bogazici University, Istanbul, Turkey (e-mail: {ilker, alagoz, ersoy}@boun.edu.tr).

H. Deliç is with the Dept. of Electrical and Electronics Engineering, Bogazici University, Istanbul, Turkey (e-mail: delic@boun.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2006.01005.

II. Coverage Model

For event-driven sensor networks, the packet traffic generated depends primarily on (i) the degree of coverage at the event point, which is defined as the number of sensor nodes that sense the event point, and (ii) the distribution of the events in the surveillance area. In this work, we assume that at most one target crosses the border at any given time, by which the complexity of the second factor is reduced.

Because sensing is a power consuming operation, it generally has its own duty cycle (for instance 1%, which corresponds to 10 msec sensing per second). Let t_s denote the sensing period, which means that each node senses the environment once in t_s seconds. Hence, after the target is detected by a sensor at location (x, y) at time t, it will possibly be sensed by the same node again at location (x', y') at time $t+t_s$, where the Euclidian distance between (x, y) and (x', y') is $v_T t_s$, with v_T being the velocity of the target within the $(t, t + t_s)$ period. Assuming that the sensing offset of all sensors are the same, the traffic (number of data packets) generated at point (x, y) is equal to the coverage degree, $c_{x,y}$, of that location.

There are a number of sensing range models that can be employed for intrusion detection (e.g., Elfes, Neyman-Pearson [4]). For the sake of simplicity, we use binary sensing, where the target is sensed with probability one, when it is within the sensing range (R_s) of the node. Then, the coverage degree probability of a point is related to the probability of the number of sensors within the distance R_s of this point. That is because, only the nodes that are at most R_s away can sense the target. For each sensor node deployed, the probability of that node to be within the R_s-distance of the point is a Bernoulli trial, where the probability of success is

1089-7798/06$20.00 © 2006 IEEE

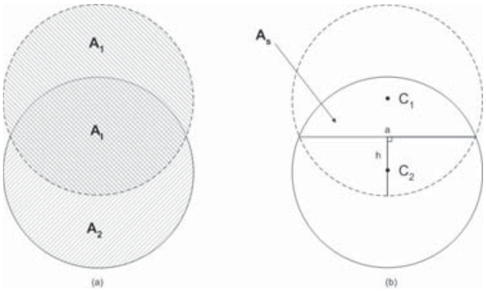

Fig. 1. Geometric representation of successive target detection.

$$p = \frac{\pi R_s^2}{LW} \quad (1)$$

where (L,W) is the length and width of the borders of the surveillance area. Hence, the number of sensor nodes within distance R_s of a point forms a Binomial distribution. Moreover, for large N and small p, which is generally the case in intrusion detection applications, this Binomial distribution can be represented by a Poisson process. Then, the mean value of the equivalent Poisson process is

$$\lambda = Np = \frac{N\pi R_s^2}{LW}. \quad (2)$$

III. Packet Traffic Model

The coverage degree probabilities of two nearby surveillance area points are not independent of each other. This suggests that the coverage degrees of all points of a surveillance area cannot be modeled with the Poisson distribution. This dependency is important since network traffic models should define a time-based traffic load. If we are given the number of sensing nodes at time t, we cannot use the Poisson distribution presented in Section II to estimate the number of nodes that will detect the target at time $t + t_s$.

The reason behind the degree-dependency is indicated in Fig. 1b. Let C_1 and C_2 denote successive target detection points. The distance between C_1 and C_2 is equal to $v_T t_s$. In addition, the coverage degree of point C_1 (C_2) equals the number of sensor nodes residing in $Circle_1$ ($Circle_2$), where $Circle_i$ is the circle whose center is at C_i and whose radius is R_s. The dependency of the coverage degrees of points C_1 and C_2 is represented by the intersection of the two circles[1].

To investigate the dependency of the coverage degrees, we have to first look into the deployment probabilities of the crescent areas A_1 and A_2, and the intersection area A_I shown in Fig. 1. Let the random variables X_i and Y_i denote the coverage degree of point C_i and the number of nodes that reside in A_i, respectively. Then,

$$P(Y_I + Y_i = n) = P(X_i = n), \quad i = 1, 2. \quad (3)$$

Given that point C_1 has coverage degree c_1, the probability that point C_2 has coverage degree c_2 is found as follows.

[1]Without loss of generality, assume that $Circle_2$ does not intersect with $Circle_0$ whose center is the detection point before C_1, which can be designated as C_0.

Define $c_{\min} = \min(c_1, c_2)$. Then,

$$P(X_2 = c_2 | X_1 = c_1) = \sum_{i=0}^{c_{\min}} P(Y_I = i | X_1 = c_1) P(X_2 = c_2 | Y_I = i). \quad (4)$$

If it is known that there exist c_1 sensors within the first circle, then the probability of having i of them inside A_I possesses the Binomial distribution, where the probability of success is $A_I / \pi R_s^2$. Hence,

$$P(Y_I = i | X_1 = c_1) = \binom{c_1}{i} \left(\frac{A_I}{\pi R_s^2}\right)^i \left(1 - \frac{A_I}{\pi R_s^2}\right)^{c_1 - i}. \quad (5)$$

Moreover, the probability of having $c_2 - i$ sensors within A_2 again possesses the Binomial distribution. However, c_1 sensors are known to be out of that area. Hence, we are left with $N - c_1$ sensors to be deployed in the entire surveillance area minus the disc contained by $Circle_1$. Therefore,

$$P(X_2 = c_2 | Y_I = i) = P(Y_2 = c_2 - i) = \binom{N - c_1}{c_2 - i} \times \left(\frac{A_2}{LW - \pi {R_s}^2}\right)^{c_2 - i} \left(1 - \frac{A_2}{LW - \pi {R_s}^2}\right)^{N - c_1 - (c_2 - i)} \quad (6)$$

An important observation of (4) is that the time-dependent random variable X_t, which represents the coverage degree probability of the target location at time t, corresponds to a Markov process. The dependency between different instances of this random variable is a result of the circle intersection described in Fig. 1. Based on the system parameters, a circle that represents a target location at any time instance may intersect with the nth previous one. Then, the Markov process is referred to be of nth order (i.e., it is dependent on the previous n states but no further back), which is calculated as:

$$n v_T t_s < 2R_s \leq (n+1) v_T t_s \quad \Rightarrow \quad n = \left\lfloor \frac{2R_s}{v_T t_s} \right\rfloor, \quad (7)$$

where $n \in \{0, 1, \ldots\}$. In the sequel, we only consider the case where X_t is a first-order Markov process.

The evaluation of the probabilities derived above requires the calculation of the intersection area of the two circles. The intersection area is equal to twice the area of the circular segment represented as A_s in Fig. 1b. The exact value of A_s is given in [3] as

$$A_s = R_s^2 \cos^{-1}\left(\frac{R_s - h}{R_s} - (R_s - h)\sqrt{2R_s h - h^2}\right). \quad (8)$$

The calculation of A_s is not trivial due to the inverse cosine function. However, using the approximation by Harris *et al.* [5],

$$A_s = \frac{2}{3} a h + \frac{h^3}{2a}, \quad (9)$$

which is found to be accurate to within 0.1% for $0^o \leq \theta \leq 150^o$ and 0.8% for $150^o \leq \theta \leq 180^o$ where θ is the central angle that covers the chord a. The lengths of the chord a and h are related with the velocity of the target v_T and the sensing interval t_s:

$$a = 2\sqrt{R_s^2 - \left(\frac{v_T t_s}{2}\right)^2}, \quad (10a)$$

$$h = R_s - \frac{v_T t_s}{2}. \quad (10b)$$

The following algorithm is used to produce sample packet traffic with the assumption that the number of packets generated at any point equals the coverage degree of that point:

Algorithm 1 Packet Traffic Generation Algorithm

1: Set c_0 to be a random value chosen from Poisson distribution with mean $\lambda = \frac{N\pi R_s^2}{LW}$ {*entrance point cov. deg.*}
2: Calculate A_s using (9) and (10).
3: **for** $t = 1$ to $\lfloor \frac{W}{v_T t_s} \rfloor$ {*assuming a shortest crossing path*} **do**
4: Choose a value for c_t randomly, based on the probabilities found in (4)-(6).
5: **end for**

IV. Experimental Results

As a reference scenario, we set the system parameter values as specified in Table I, and investigate the coverage degrees of the sensor nodes under uniformly distributed deployment. The number of nodes indicated in Table I is selected so that if regular grid deployment is employed, that many nodes are required for a minimum coverage of 99% of the surveillance area.

TABLE I
Reference Scenario

Length (L)	10000 m
Width (W)	1000 m
Number of sensors (N)	10000
Sensing Range (R_s)	20 m
Target Velocity (v_T)	10 m/sec
Sensing Intervals (t_s)	1 sec

The simulation results indicated in Fig. 2 show that the coverage degree at any point is a Poisson-distributed random variable. Different tones of lines represent 1000 different uniformly distributed deployments where coverage degrees of any 100 points of each deployment are collected.

The simulation results shown in Fig. 3 prove the dependency of the neighboring point probabilities. Neighboring points define the next potential detection points of the target, which are 10 m away for the scenario depicted by Table I ($v_T t_s$). It is clear that the coverage degree of next potential detection point is not Poisson-distributed.

Based on (4)-(6), simulation results for the neighbors of the points with a degree of two are compared with the analytic calculations for the reference scenario. As seen in Fig. 3, the analysis made for the coverage degree calculations can be used to model the coverage of the subsequent points where the target is sensed, which can be used to model the offered WSN traffic.

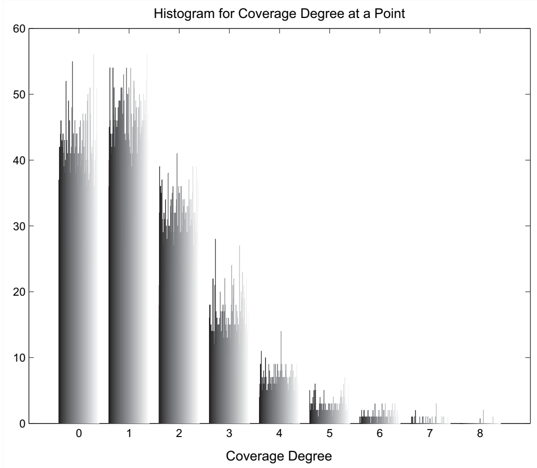

Fig. 2. Coverage degree histogram.

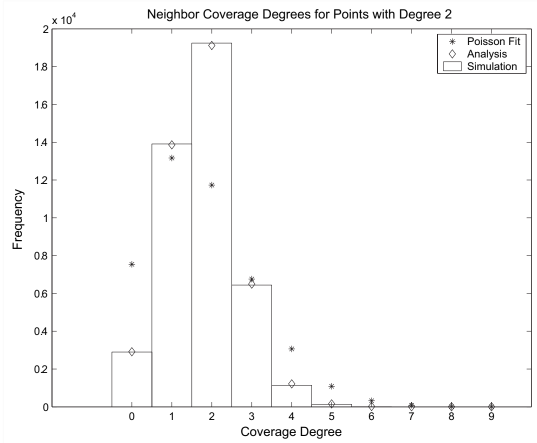

Fig. 3. Neighbor coverage degree histogram for points with degree two.

V. Conclusion and Future Work

In this letter, a new packet traffic model is devised for intrusion detection applications. The system design parameters considered in this model are the number of sensor nodes deployed, surveillance area, sensing range, target velocity and sampling interval. Simulation results support the analytical work presented. We plan to extend the analysis to lift certain assumptions such as uniformly distributed deployment, single target, binary sensing and single data packet generation per target sensing.

References

[1] S. Cui, R. Madan, A. J. Goldsmith, and S. Lall, "Joint routing, MAC, and link layer optimization in sensor networks with energy constraints," in *Proc. IEEE ICC 2005*, vol. 2, pp.725-729, May 2005.
[2] Y. Ma and J. H. Aylor, "System lifetime optimization for heterogeneous sensor networks with a hub-spoke topology," *IEEE Trans. Mobile Computing*, vol. 3, pp. 286-294, July-Sept. 2004.
[3] H. Takagi and L. Kleinrock, "Optimal transmission ranges for randomly distributed packet radio terminals," *IEEE Trans. Commun.*, vol. 32, pp. 246-257, Mar. 1984.
[4] E. Onur, C. Ersoy, and H. Deliç, "How many sensors for an acceptable breach probability level?" *Computer Communications* (Special Issue on Dependable Wireless Sensor Networks), in production, 2005.
[5] J. W. Harris and H. Stocker, "Segment of a Circle," in *Handbook of Mathematics and Computational Science*, pp. 92-93. New York: Springer-Verlag, 1998.